

The post-genomics era of cotton

Yuxian Zhu

Institute for Advanced Studies and College of Life Science, Wuhan University, Wuhan 430072, China

Received September 11, 2015; accepted September 12, 2015; published online January 22, 2016

Citation: Zhu, Y. (2016). The post-genomics era of cotton. *Sci China Life Sci* 59, 109–111. doi: 10.1007/s11427-016-5017-6

As one of the important industrial crops and also an excellent model organism for studying cell elongation, cell wall biosynthesis as well as polyploidy evolution, cotton has always been in the limelight of the world (Shi et al., 2006; Qin and Zhu, 2011). With the publication of the allotetraploid *Gossypium hirsutum* and *Gossypium barbadense* genomes this year (Li et al., 2015; Zhang et al., 2015; Liu et al., 2015), all three important cotton genomes, including the two ancestor diploid *Gossypium raimondii* (DD) (Wang et al., 2012; Paterson et al., 2012) and *Gossypium arboreum* (AA) (Li et al., 2014) have been successfully sequenced by scientists from China and from the United States of America. Since functional studies of *Arabidopsis thaliana* and *Oryza sativa* have largely been promoted by their respective genome sequencing projects (Goff et al., 2002; Initiative TAG, 2000; Yu et al., 2002), we believe that the sequenced cotton genomes will ultimately stimulate fundamental researches on genome evolution, gene mapping, gene cloning, transcriptomics, epigenomics, proteomics and applied research such as molecular breeding. The post-genomics era of cotton research has finally come into sight.

It is my great honor to guest-edit this thematic issue on cotton biology in *Science China Life Sciences*. In this issue, I am pleased to present an epitome on recent cotton researches derived from quite many important cotton research laboratories in China. One group each of scientists analyzed the phylogeny, chromosomal location, conserved domains and gene structure as well as the expression patterns of the phospholipase D gene family in *G. arboreum* (Tang et al., 2016) and the NAC transcription factor (NAC-TF) family in

both diploid cottons (Shang et al., 2016). A total of 19 non-redundant *PLD* genes (*GaPLDs*), belonging to six well-supported clades (α , β/γ , δ , ϵ , ζ and ϕ), were identified. Most of the *GaPLD* genes within the same clade showed similar exon-intron organization with highly conserved motif structures. Segmental duplication that was estimated to occur from 19.61 to 20.44 million years ago was found the major contributor to the expansion of the *GaPLD* gene family in cotton (Tang et al., 2016). Genome-wide annotation identified 145 and 141 NAC genes, respectively. Genes from both diploid cotton genomes were divided into 18 subfamilies with distinct gene structure divergence. Analysis of expression patterns of homologous gene pairs suggested that diversification of gene function has taken place during the evolution of *G. raimondii* and *G. arboreum* (Shang et al., 2016).

Jin et al. performed global 2-DE analysis in Upland cotton to provide the largest protein datasets from ovules and developing fibers (Jin et al., 2016). They collected 1,203 independent protein spots from representative 2-DE gels with 975 of them successfully identified by MALDI TOF-TOF. Functional annotation revealed that the ovule preferred proteins were mainly enriched in fatty acid elongation, sulfur amino acid metabolism and postreplication repair, while fiber preferred proteins were enriched in root hair elongation pathway, galactose metabolism and D-xylose metabolism.

Our group analyzed the transposable elements (TEs), or the so-called ‘junk DNA’, mediated gene regulations and proposed a potential role for TEs during cotton genome evolution. TEs accounted for up to 57%, 68.5% and 67.2%, respectively in DD, AA and AtAtDtDt genomes. The

email: zhuyx@whu.edu.cn

1,694-Mb AA-genome was found to harbor more LTR-type retrotransposons that made cardinal contributions to the twofold increase in its genome size after evolution from the 775.2-Mb DD-genome. Although the 2,173 Mb AtDt genome showed similar TE content to the A-genome, the total numbers of *LTR-Gypsy* and *LTR-Copia* type TEs varied significantly between these two genomes, indicating that TEs may also play some role during cotton fiber development (Wang et al., 2016a). Using strand-specific RNA sequencing (ssRNA-seq) technology with samples either from cotton fibers or leaves, Zou et al. identified 5,996 long noncoding RNAs (lncRNAs), including 3,510 long intergenic noncoding RNAs (lincRNAs) and 2,486 natural antisense transcripts (lncNAT) from *G. arborum*. Expression analysis revealed that 51.9% of lincRNAs and 54.5% of lncNATs transcripts were preferentially expressed at one or more stages of fiber development. This is significantly higher than the rate of preferentially expressed protein-coding transcripts which was about 21.7% (Zou et al., 2016).

Wang et al. tested PNZIP, a green tissue-specific promoter to drive Bt expression in cotton which show strong toxicity to both cotton bollworm and pink bollworm with very low toxin accumulation in seeds. They cloned the PNZIP (*pharbitis nil* leucine zipper) gene promoter, fused it with both the *GUS* and the *Cry 9C* gene and obtained several transgenic cotton lines via *Agrobacterium*-mediated transformation. Comparing to that of the transgenic CaMV 35S::Cry 9C plants, in seeds of some PNZIP::Cry 9C transgenic line, the amount of Cry 9C protein was reduced to less than 1% whereas relatively similar levels of the protein were found in various vegetative tissues, including rinds of the green ball and bracts. This strategy will likely liberate the public's concern about food safety when consume transgenic cotton seed oil (Wang et al., 2016b).

Being at this post-genomic era, we witness sequencing technology improvement almost in each single passing day. Great progresses in cotton research in China are expected in the following area: (i) pan-genome analysis to uncover cotton evolution, diversification and introduction of important agronomical traits. (ii) Fine-mapping and cloning of functional cotton genes by deep sequencing techniques. (iii) Integration of various 'omics' approaches to elucidate the paths of cotton gene expression and regulation during development or under environment stresses. (iv) Molecular breeding via genome editing. We have sufficient reasons to anticipate an exciting and fruitful future for cotton research.

Goff, S.A., Riche, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R.,

Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa* L. Ssp. Japonica). *Science* 296, 92–100.

Initiative TAG. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

Jin, X., Wang, L., He, L., Feng, W., and Wang, X. (2016). Two dimensional electrophoresis gel based analysis provide global insights of cotton ovule and fiber proteomes. *Sci China Life Sci* 59, 154–163.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R.G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., Ye, W., Liu, X., Zhang, X., Liu, W., Wei, H., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J., Zhu, Y., and Yu, S. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* tm-1) provides insights into genome evolution. *Nat Biotech* 33, 524–530.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., Liu, W., Shi, C., Xiao, G., Gou, C., Ye, W., Xu, X., Zhang, X., Wei, H., Li, Z., Zhang, G., Wang, J., Liu, K., Kohel, R.J., Percy, R.G., Yu, J.Z., Zhu, Y.X., Wang, J., and Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet* 46, 567–572.

Liu, X., Zhao, B., Zheng, H., Hu, Y., Lu, G., Yang, C., Chen, J., Chen, J., Chen, D., Zhang, L., Zhou, Y., Wang, L., Guo, W., Bai, Y., Ruan, J., Shanguan, X., Mao, Y., Shan, C., Jiang, J., Zhu, Y., Jin, L., Kang, H., Chen, S., He, X., Wang, R., Wang, Y., Chen, J., Wang, L., Yu, S., Wang, B., Wei, J., Song, S., Lu, X., Gao, Z., Gu, W., Deng, X., Ma, D., Wang, S., Liang, W., Fang, L., Cai, C., Zhu, X., Zhou, B., Jeffrey, C., Xu, S., Zhang, Y., Wang, S., Zhang, T., Zhao, G., and Chen, X. (2015). *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci Rep* 5, 14139.

Paterson, A.H., Wendel, J.F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K.C., Shu, S., Udall, J., Yoo, M.J., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M.V., Gong, L., Grimwood, J., Grover, C., Grupp, K., Hu, G., Lee, T.H., Li, J., Lin, L., Liu, T., Marler, B.S., Page, J.T., Roberts, A.W., Romanel, E., Sanders, W.S., Szadkowski, E., Tan, X., Tang, H., Xu, C., Wang, J., Wang, Z., Zhang, D., Zhang, L., Ashrafi, H., Bedon, F., Bowers, J.E., Brubaker, C.L., Chee, P.W., Das, S., Gingle, A.R., Haigler, C.H., Harker, D., Hoffmann, L.V., Hovav, R., Jones, D.C., Lemke, C., Mansoor, S., Rahman, M.U., Rainville, L.N., Rambani, A., Reddy, U.K., Rong, J.K., Saranga, Y., Scheffler, B.E., Scheffler, J.A., Stelly, D.M., Triplett, B.A., Van Deynze, A., Vaslin, M.F.S., Waghmare, V.N., Walford, S.A., Wright, R.J., Zaki, E.A., Zhang, T., Dennis, E.S., Mayer, K.F.X., Peterson, D.G., Rokhsar, D.S., Wang, X., and Schmutz, J. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427.

Qin, Y., and Zhu, Y. (2011). How cotton fibers elongate: A tale of linear cell-growth mode. *Curr Opin Plant Biol* 14, 106–111.

Shang, H., Wang, Z., Zou, C., Zhang, Z., Li, W., Li, J., Shi, Y., Gong, W., Chen, T., Liu, A., Gong, J., Ge, Q., and Yuan, Y. (2016). Comprehensive analysis of NAC transcription factors in diploid *Gossypium*: Sequence conservation and expression analysis uncover their roles during fiber development. *Sci China Life Sci* 59, 142–153.

Tang, K., Dong, C., and Liu, J. (2016). Genome-wide analysis and expression profiling of the phospholipase D gene family in *Gossypium arboreum*. *Sci China Life Sci* 59, 130–141.

Shi, Y., Zhu, S., Mao, X., Feng, J., Qin, Y., Zhang, L., Cheng, J., Wei, L., Wang, Z., and Zhu, Y. (2006). Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18, 651–664.

Wang, K., Huang, G., and Zhu, Y. (2016a). Transposable elements play an important role during cotton genome evolution and fiber cell development. *Sci China Life Sci* 59, 112–121.

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., Zou, C., Li, Q., Yuan, Y., Lu, C., Wei, H., Gou, C., Zheng, Z., Yin, Y., Zhang, X., Liu, K., Wang, B., Song, C., Shi, N.,

- Kohel, R.J., Percy, R.G., Yu, J.Z., Zhu, Y., Wang, J., and Yu, S. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* 44, 1098–1103.
- Wang, Q., Zhu, Y., Sun, L., Li, L., Jin, S., and Zhang, X. (2016b). Transgenic Bt cotton driven by the green tissue-specific promoter shows strong toxicity to lepidopteran pests and lower BT toxin accumulation in seeds. *Sci China Life Sci* 59, 172–182.
- Yu, J., Hu, S., Wang, J., Wong, G., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., and Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. Ssp. Indica). *Science* 296, 79–92.
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., Hulse-Kemp, A.M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., Chang, L., Zhang, W., Song, Q., Kirkbride, R.C., Chen, X., Dennis, E., Llewellyn, D.J., Peterson, D.G., Thaxton, P., Jones, D.C., Wang, Q., Xu, X., Zhang, H., Wu, H., Zhou, L., Mei, G., Chen, S., Tian, Y., Xiang, D., Li, X., Ding, J., Zuo, Q., Tao, L., Liu, Y., Li, J., Lin, Y., Hui, Y., Cao, Z., Cai, C., Zhu, X., Jiang, Z., Zhou, B., Guo, W., Li, R., and Chen, Z. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. Acc. Tm-1) provides a resource for fiber improvement. *Nat Biotech* 33, 531–537.
- Zou, C., Lu, C., Wang, Q., Lu, C., Yang, W., Zhang, Y., Cheng, H., Feng, X., Prosper, M.A., and Song, G. (2016). Transcriptome analysis reveals long noncoding RNAs involved in fiber development of cotton (*Gossypium arboreum*). *Sci China Life Sci* 59, 164–171.



Biographical Sketch

Yuxian Zhu, a well-known Plant Physiologist in China. He was born in Fuyang, Zhejiang Province, December 1955. He is a member of the Chinese Academy of Sciences and also a member of The World Academy of Sciences for developing countries. Currently, he is the Dean of the Institute for Advanced Studies at Wuhan University. His research is mainly focused on plant molecular biology, especially on gene regulations during cotton fiber development as well as cotton genomics.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.